

1. Analisi esplorative

1.1 Analisi Univariata

La Hydronium concentration è l'operazione inversa per trovare il pH, ed è la concentrazione di idronio

Il database è composto da 44 osservazioni, una per ciascun fiume, e 9 variabili. Le variabili rappresentano varie caratteristiche dei fiumi come il numero di specie di cozze presenti, la dimensione del fiume, la distanza da alcuni fiumi importanti, e vari indicatori sulla qualità delle acque.

	Min.	1 ^o qt.	Mediana(IQR)	Media(sd)	3 ^o qt.	Max
Species	2.00	8.00	10.00(5.0)	11.25(5.99)	13.00	33.00
Area	349	2115	4315(7797.5)	6590(6016)	9912	27900
Stepping stones to AC	1.00	7.00	15.50(15.25)	15.34(9.19)	22.25	33.00
Stepping stones to AP	0.00	4.00	12.00(14.25)	12.02(8.24)	18.25	28.00
Stepping stones to SV	0.00	5.00	7.00(13.25)	8.136(8.94)	11.00	21.00
Stepping stones to SL	4.00	16.75	22.00(0.775)	22.16(1.84)	30.00	36.00
Nitrate	0.100	0.600	0.800(0.775)	1.495(1.84)	1.375	8.700
Solid Residue	29.0	56.5	78.0(64.0)	112.4(17.3)	120.5	520.0
Hydronium	0.200	1.00	1.600(2.2)	3.631(6.03)	3.200	32.00
ln(Area)	5.855	7.657	8.370(1.54)	8.331(1.07)	9.201	10.24

Tabella 1: Tabella di sintesi

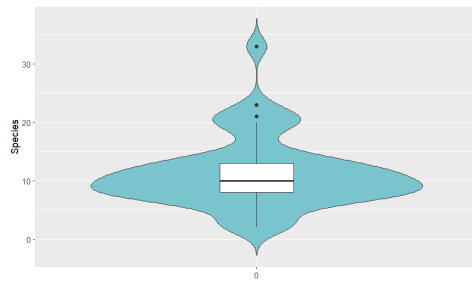


Figura 1: Violinplot della variabile
Species

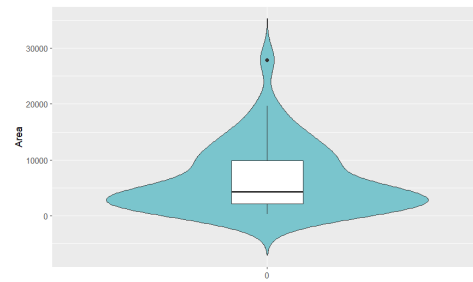


Figura 2: Violinplot della variabile
Area

La variabile **Species**, rappresentante il numero di specie di cozze nel fiume, ha una distribuzione asimmetrica a destra come evidenziato dal violinplot (*Figura 1*). Ha un valore minimo di 2 specie (nel fiume Waccasassa) e un massimo di 33 specie (nel fiume Apalachicola). La mediana è di 10 specie, la media è di 11.25 specie. La variabile **Area** segue una distribuzione asimmetrica a destra (*Figura 2*), con un valore minimo di 349 sq mi (*square mile*) e un massimo di 27900 sq mi. La mediana è di 4315 sq mi, la media è di 6590 sq mi (*Tabella 1*).

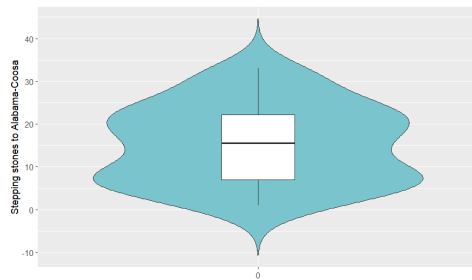


Figura 3: Violinplot della variabile
Stepping stones to AC

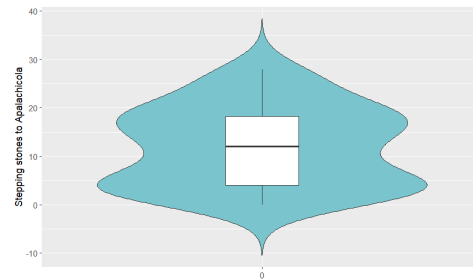


Figura 4: Violinplot della variabile
Stepping stones to AP

La distanza dal sistema fluviale di Alabama-Coosa (**Stepping stones to AC**) segue una distribuzione simmetrica (*Figura 3*), suggerendo che i fiumi sono equamente distribuiti rispetto a questo sistema fluviale. La distanza minima è di 1 e la massima di 33. La distanza media è di 15.34, e quella mediana di 15.50 "saltelli". Anche la distanza dal fiume Apalachicola (**Stepping stones to AP**) segue una distribuzione simmetrica (*Figura 4*), con una distanza minima di 0 (coincidente con il fiume stesso) e una massima di 28. La distanza media è di 12.02, e quella mediana di 12.00 "saltelli" (*Tabella 1*).

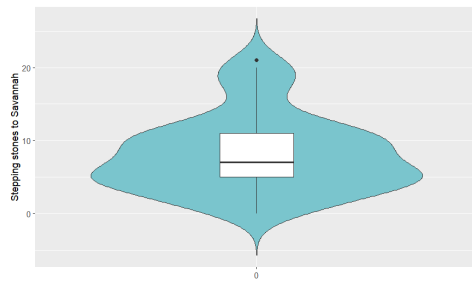


Figura 5: Violinplot della variabile
Stepping stones to SV

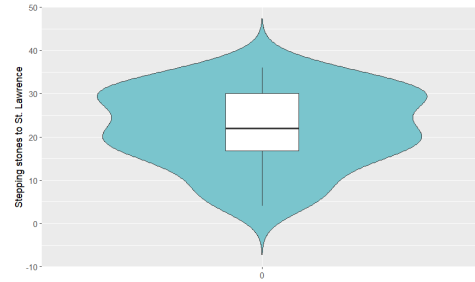


Figura 6: Violinplot della variabile
Stepping stones to SL

La distanza dal fiume Savannah (**Stepping stones to SV**) ha una distribuzione simmetrica anche se presenta una leggera coda lunga a destra (*Figura 5*), con una distanza minima di 0 (coincidente con il fiume stesso) e una massima di 21. La distanza media è di 8.136, e quella mediana di 7.00. Il valore molto basso della mediana e la presenza di una piccola coda lunga a destra suggeriscono che la maggior parte dei fiumi osservati sono molto vicini al fiume Savannah. La distanza dal fiume St. Lawrence (**Stepping stones to SL**) ha una distribuzione simmetrica (*Figura 6*), ma presenta anche una leggera coda lunga a sinistra. La distanza minima di 4 e una massima di 36. La distanza media è di 22.16, e quella mediana di 22.00. Il valore elevato della di media e mediana, in aggiunta alla leggera coda a sinistra, suggerisce che la maggior parte dei fiumi sono distanti al fiume St. Lawrence (*Tabella 1*).

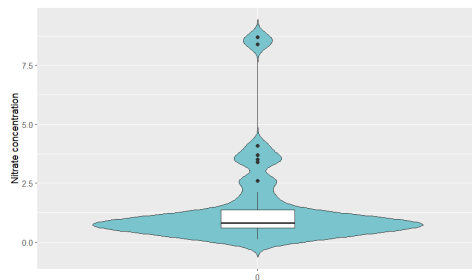


Figura 7: Violinplot della variabile
Nitrate

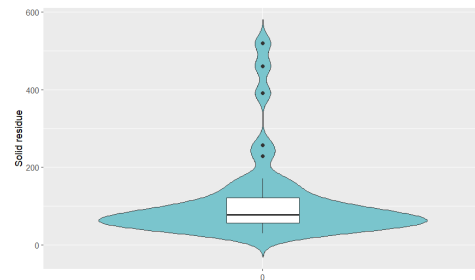


Figura 8: Violinplot della variabile
Solid Residue

La variabile **Nitrate** ha una distribuzione asimmetrica a destra (*Figura 7*), con un valore minimo di 0.100 ppm (*parts per million*) e un massimo di 8.700 ppm. La mediana è di 0.800 ppm, la media è di 1.495 ppm. la variabile **Solid Residue** ha anche essa una distribuzione asimmetrica a destra (*Figura 8*), con un valore minimo di 29.0 ppm e un massimo di 520.0 ppm. La mediana è di 78.0 ppm, la media è di 112.4 ppm (*Tabella 1*).

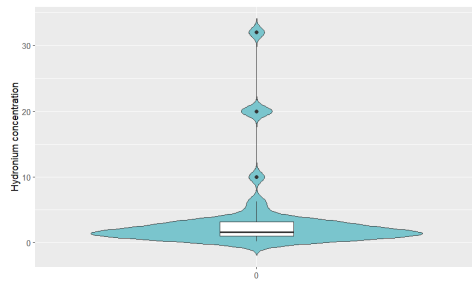


Figura 9: Violinplot della variabile
Hydronium

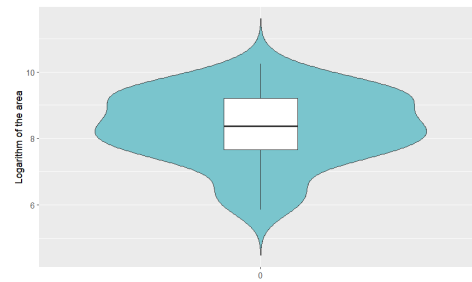


Figura 10: Violinplot della variabile
 $\ln(\text{Area})$

La variabile **Hydronium** ha una distribuzione asimmetrica a destra (*Figura 9*), con un valore minimo di 0.200 e un massimo di 32.00. La mediana è di 1.600, la media è di 3.631. La variabile **$\ln(\text{Area})$** segue una distribuzione simmetrica (*Figura 10*), con un valore minimo di 5.855 e un massimo di 10.24. La mediana è di 8.370, la media è di 8.331 (*Tabella 1*).

1.2 Analisi Bivariate

Si valuta la relazione tra le varie variabili. Delle 36 bivariate totali, 16 bivariate sono risultate significative (*Tabella 2*).

	ln(A)	H ⁺	SR	NO ₃	SL	SV	AP	AC
Sp	1	0	0	0	0	0	0	0
AC	0	1	0	1	1	1	1	
AP	0	1	0	1	1	1		
SV	0	0	0	1	1			
SL	0	1	0	1				
NO ₃	0	0	1					
SR	0	1						
H ⁺	0							

Tabella 2: Tabella che indica la significatività delle bivariate, 1 indica una bivariata significativa, 0 indica una bivariata non significativa, In particolare Sp: Species, AC: Stepping stones to AC, AP: Stepping stones to AP, SV: Stepping stones to SV, SL: Stepping stones to SL, NO₃: Nitrate, SR: Solid Residue, H⁺: Hydronium, ln(A): ln(Area)

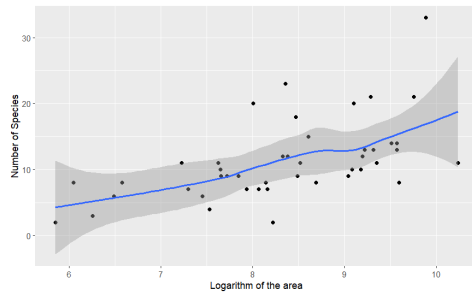


Figura 11: Diagramma di dispersione tra Species e ln(Area), con curva LOESS (*LOcally Estimated Scatterplot Smoothing*) in blu e intervallo di confidenza al 95% in grigio

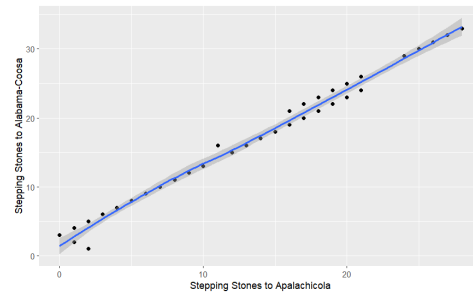


Figura 12: Diagramma di dispersione tra Stepping stones to AC e Stepping stones to AP, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra il numero di specie di cozze e il logaritmo dell'area (*Figura 11*) si nota una relazione positiva tra le due variabili. La correlazione di Spearman è 0.64, indicando una relazione positiva tra le due variabili; dunque all'aumentare del logaritmo dell'area del fiume aumenta anche il numero di specie

di cozze presenti. Il test per verificare se la correlazione è significativa ($S=5089.2$, $p\text{-value}=0.00001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa che conferma la presenza di una relazione positiva tra le due variabili.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal sistema fluviale Alabama-Coose e dal fiume Apalachicola (*Figura 12*) si nota una relazione positiva, quasi lineare, tra le due variabili. La correlazione di Spearman è 0.99, indicando una forte relazione positiva tra le due variabili; dunque all'aumentare della distanza di un fiume dal sistema fluviale Alabama-Coose aumenta anche la distanza dal fiume Apalachicola. Questa relazione è dovuta al fatto che i due fiumi sono molto vicini tra loro, scorrono in parallelo, e sfociano entrambi nel Golfo del Messico, e quindi la distanza da uno implica la distanza dall'altro. Il test per verificare se la correlazione è significativa ($S=89.42$, $p\text{-value}=0.00001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

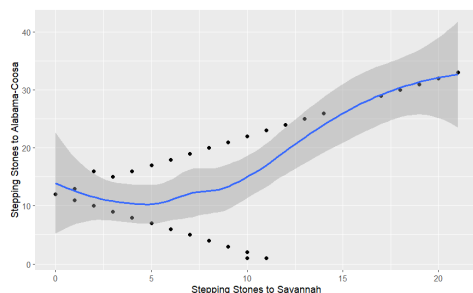


Figura 13: Diagramma di dispersione tra Stepping stones to AC e Stepping stones to SV, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

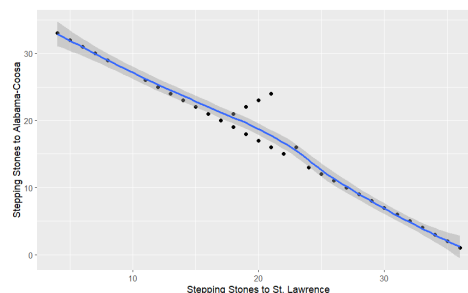


Figura 14: Diagramma di dispersione tra Stepping stones to AC e Stepping stones to SL, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal sistema fluviale Alabama-Coose e dal fiume Savannah (*Figura 13*) si nota una relazione tendenzialmente positiva, ma molto curvilinea tra le due variabili. In particolare si nota come nella prima del grafico i punti divergono tra di loro. Questo è dovuto al fatto che il fiume Savannah e il sistema fluviale Alabama-Coose scorrono perpendicolarmente tra di loro, quasi a formare un angolo retto. Quindi una porzione di fiumi (quelli situati a Sud), mentre si allontanano dal fiume Savannah si avvicinano al sistema fluviale Alabama-Coose. Al contrario, i fiumi situati a Nord si allontanano parallelamente ad entrambi i fiumi che spiega la crescita positiva nella seconda parte del grafico. La correlazione di Spearman è 0.54, indicando una modesta relazione positiva tra le due variabili. La correlazione è positiva; dunque all'aumentare della distanza di un fiume dal sistema fluviale Alabama-Coose aumenta anche la distanza dal fiume Savannah. Il test per verificare se la correlazione è significativa ($S=6483.6$, $p\text{-value}=0.0001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal sistema fluviale Alabama-Coose e dal fiume Saint Lawrence (*Figura 14*) si nota una relazione negativa, abbastanza lineare, tra le due variabili. La correlazione di Spearman è -0.97, indicando una forte relazione negativa tra le due variabili; dunque all'aumentare della distanza di un fiume dal sistema fluviale Alabama-Coose diminuisce la distanza dal fiume St. Lawrence. Questo è dovuto dalla posizione dei due fiumi. Essi, infatti, si trovano alle due estremità della regione osservata, con il fiume St. Lawrence situato più a Nord-Est e il sistema fluviale Alabama-Coose situato invece a Sud-Ovest. Quindi allontanandosi dal sistema fluviale Alabama-Coose ci si avvicina al fiume St. Lawrence, e viceversa. Il test per verificare se la correlazione è significativa ($S=28075$, $p\text{-value}=0.000001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

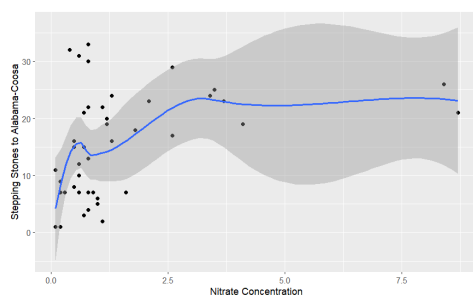


Figura 15: Diagramma di dispersione tra **Stepping stones to AC** e **Nitrate**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

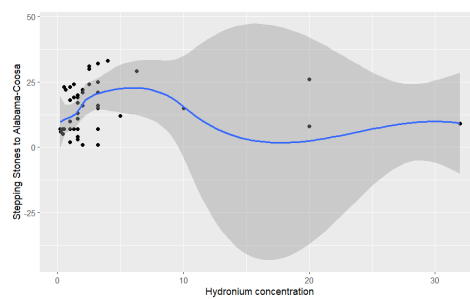


Figura 16: Diagramma di dispersione tra **Stepping stones to AC** e **Hydronium**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal sistema fluviale Alabama-Coose e la concentrazione di nitrato nelle acque dei fiumi osservati (*Figura 15*) si osserva una relazione molto curvilinea. In particolare si nota come i fiumi più vicini al sistema fluviale Alabama-Coose presentano una minore concentrazione di nitrato. La concentrazione di nitrato aumenta all'aumentare della distanza dal sistema fluviale Alabama-Coose, ma solo fino ad un certo punto. Infatti, dopo una certa distanza, la concentrazione di nitrato inizia a diminuire. La correlazione di Spearman è 0.42, indicando una lieve relazione positiva tra le due variabili. Il test per verificare se la correlazione è significativa ($S=8173.9$, $p\text{-value}=0.004$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal sistema fluviale Alabama-Coose e la variabile **Hydronium** (*Figura 16*) si nota come la maggior parte dei fiumi presenti valori bassi di concentrazione di idronio. Fanno eccezione 5 fiumi che invece presentano valori molto alti di concentrazione di idronio. La correlazione di Spearman è 0.33, indicando una lieve relazione positiva tra le due variabili.

Il test per verificare se la correlazione è significativa ($S=9506.5$, $p\text{-value}=0.029$) porta a rifiutare l'ipotesi che la correlazione non sia significativa al 5%, ma non all'1%.

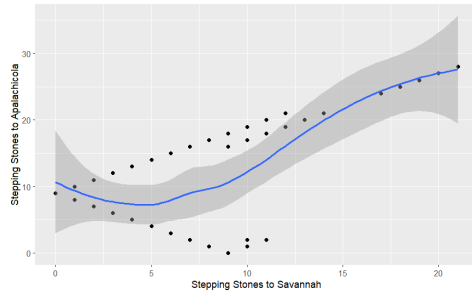


Figura 17: Diagramma di dispersione tra **Stepping stones to AP** e **Stepping stones to SV**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

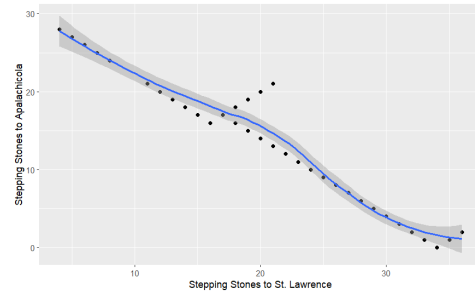


Figura 18: Diagramma di dispersione tra **Stepping stones to AP** e **Stepping stones to SL**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume Apalachicola e dal fiume Savannah (*Figura 17*) si nota una relazione tendenzialmente positiva, ma molto curvilinea tra le due variabili. Il grafico mostra una situazione simile a quella osservata tra la distanza dei fiumi dal sistema fluviale Alabama-Coose e dal fiume Savannah in *Figura 13*. Anche in questo caso, infatti, si osserva nella prima parte un divergere delle distanze e successivamente un crescere lineare delle stesse. Ciò è dovuto alla posizione del fiume Apalachicola che è molto vicino al sistema fluviale Alabama-Coose e quindi questo causa un comportamento simile. La correlazione di Spearman è 0.55, indicando una modesta relazione positiva tra le due variabili. Il test per verificare se la correlazione è significativa ($S=6436.4$, $p\text{-value}=0.0001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume Apalachicola e dal fiume Saint Lawrence (*Figura 18*) si evince una situazione analoga a quella discussa precedentemente. Si nota una relazione negativa, abbastanza lineare, tra le due variabili molto simile a quella in *Figura 14*. La correlazione di Spearman è -0.97, indicando una forte relazione negativa tra le due variabili; dunque all'aumentare della distanza di un fiume dal sistema fluviale Apalachicola diminuisce la distanza dal fiume St. Lawrence. Il test per verificare se la correlazione è significativa ($S=27892$, $p\text{-value}=0.000001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

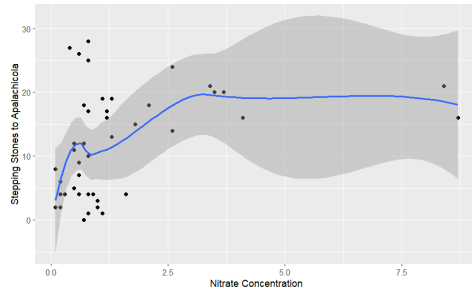


Figura 19: Diagramma di dispersione tra **Stepping stones to AP** e **Nitrate**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

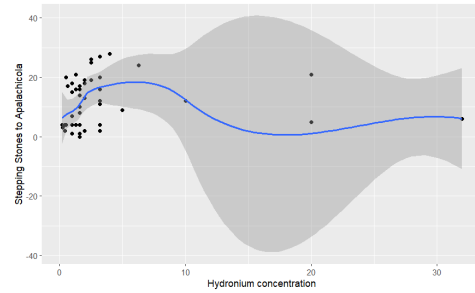


Figura 20: Diagramma di dispersione tra **Stepping stones to AP** e **Hydronium**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume Apalachicola e la concentrazione di nitrato nelle acque dei fiumi osservati (*Figura 19*) si nota come i fiumi più vicini al fiume Apalachicola presentano una bassa concentrazione di nitrato, seguita da un aumento di essa mentre ci si allontana dal fiume e infine una diminuzione della concentrazione di nitrato. La correlazione di Spearman è 0.40, indicando una lieve relazione positiva tra le due variabili. Il test per verificare se la correlazione è significativa ($S=8454.7$, $p\text{-value}=0.006$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume Apalachicola e la concentrazione di idronio (*Figura 20*) si nota come la maggior parte dei fiumi sia schiacciata nella prima parte del grafico, con valori bassi di concentrazione di idronio. Solo 5 fiumi, presentano valori molto alti di concentrazione di idronio. La correlazione di Spearman è 0.33, indicando una lieve relazione positiva tra le due variabili. Il test per verificare se la correlazione è significativa ($S=9438.8$, $p\text{-value}=0.026$) porta a rifiutare l'ipotesi che la correlazione non sia significativa al 5%.

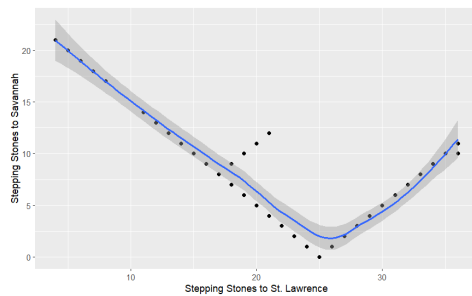


Figura 21: Diagramma di dispersione tra **Stepping stones to SV** e **Stepping stones to SL**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

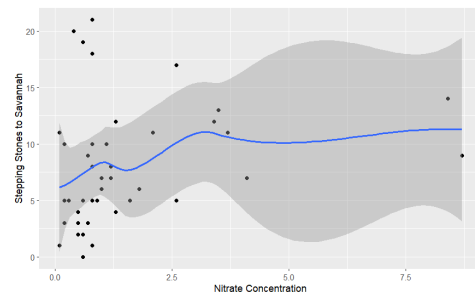


Figura 22: Diagramma di dispersione tra **Stepping stones to SV** e **Nitrate**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume Savannah e dal fiume St. Lawrence (*Figura 21*) si nota una relazione inizialmente negativa che, una volta raggiunto il fiume Savannah, cambia drasticamente e cresce positivamente. Questo andamento è dovuto alla posizione dei due fiumi. Entrambi i fiumi sfociano nell'Atlantico; il fiume St. Lawrence nasce più a Nord, nella regione dei Grandi Laghi e sfocia in Canada, il fiume Savannah invece è situato più a Sud, in Carolina del Sud. La relazione è inizialmente negativa poiché allontanandosi dal fiume St. Lawrence, e procedendo verso Sud, ci si avvicina al fiume Savannah. Una volta raggiunto il fiume Savannah, le distanze tra i due fiumi aumenteranno in parallelo. La correlazione di Spearman è -0.51, indicando una relazione negativa tra le due variabili. Il test per verificare se la correlazione è significativa ($S=21406$, $p\text{-value}=0.0004$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume Savannah e la concentrazione di nitrato nelle acque dei fiumi osservati (*Figura 22*) si nota un andamento simile a quello della *Figura 15* e della *Figura 19*, anche se la curva LOESS appare meno definita. Come nella figure precedenti, i fiumi più vicini al fiume Savannah presentano una bassa concentrazione di nitrato, seguita da un aumento di essa mentre ci si allontana dal fiume e infine una diminuzione della concentrazione di nitrato. La correlazione di Spearman è 0.36, di poco inferiore a quella tra **Stepping stones to AC** e **Nitrate Stepping stones to AP** e **Nitrate** rispettivamente di 0.42 e 0.40. La relazione è lieve ed è positiva. Il test per verificare se la correlazione è significativa ($S=9025.2$, $p\text{-value}=0.015$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

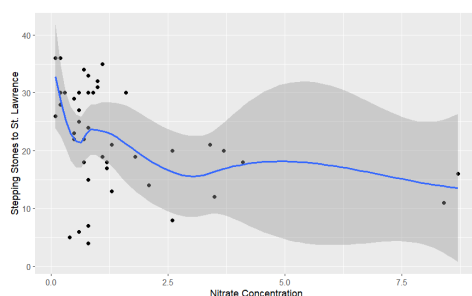


Figura 23: Diagramma di dispersione tra **Stepping stones to SL** e **Nitrate**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

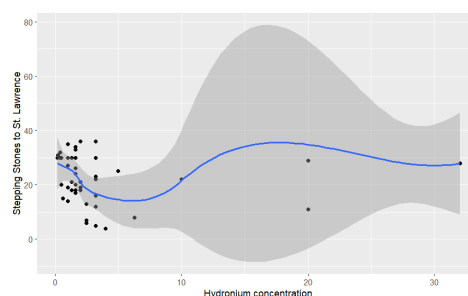


Figura 24: Diagramma di dispersione tra **Stepping stones to SL** e **Hydronium**, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

IL diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume St. Lawrence e la concentrazione di nitrato nelle acque dei fiumi osservati (*Figura 23*) mostra una situazione capovolta rispetto a quella osservata nelle *Figura 15*, *Figura 19*, *Figura 22*. Il capovolgimento è dovuto al fatto che il fiume St. Lawrence è situato più

a Nord-Est rispetto agli altri fiumi osservati. Quindi i valori che si trovavano distanti dagli altri fiumi, ora sono vicini al fiume St. Lawrence e viceversa. La correlazione di Spearman è -0.41. La correlazione è negativa, ma al netto del segno, è in linea con le correlazioni tra gli altri fiumi e la concentrazione di nitrato. Il test per verificare se la correlazione è significativa ($S=20062$, $p\text{-value}=0.005$) porta a rifiutare l'ipotesi che la correlazione non sia significativa.

Dal diagramma di dispersione tra la distanza dei vari fiumi osservati dal fiume St. Lawrence e la concentrazione di idronio (*Figura 24*) si nota come la situazione, analogamente alla *Figura 23*, sia capovolta rispetto a quella osservata nelle *Figura 16* e *Figura 20*. La correlazione di Spearman è -0.35, indicando una lieve relazione negativa tra le due variabili. Il test per verificare se la correlazione è significativa ($S=19135$, $p\text{-value}=0.02$) porta a rifiutare l'ipotesi che la correlazione non sia significativa al 5%.

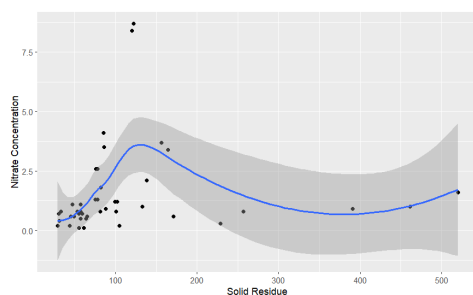


Figura 25: Diagramma di dispersione tra Nitrate e Solid Residue, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

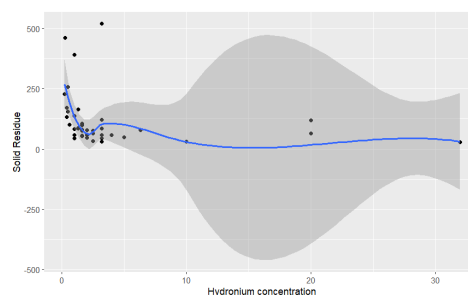


Figura 26: Diagramma di dispersione tra Solid Residue e Hydronium, con curva LOESS in blu e intervallo di confidenza al 95% in grigio

Dal diagramma di dispersione tra la concentrazione di nitrato e la concentrazione di residuo solido nelle acque dei fiumi osservati (*Figura 25*) si osserva un iniziale aumento della concentrazione di idronio all'aumentare della concentrazione di residuo solido. Tuttavia, dopo un certo punto (intorno al valore 15 per il residuo solido), la concentrazione di nitrato inizia a diminuire. La correlazione di Spearman è 0.50, indicando una modesta relazione positiva tra le due variabili. Il test per verificare se la correlazione è significativa ($S=7386.3$, $p\text{-value}=0.001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa al 5%, ma non all'1%.

Dal diagramma di dispersione tra la concentrazione di residuo solido e la variabile Hydronium (*Figura 26*) si nota una relazione molto ondulatoria tra le due variabili. La correlazione è nel complesso (tranne in una piccola parte del grafico tra i valori 2 e 3) negativa. La correlazione di Spearman è -0.55, confermando la presenza di una modesta relazione negativa tra le due variabili. Dunque, all'aumentare della concentrazione di residuo solido diminuisce la concentrazione di idronio. Questo andamento

ha molto senso in quanto il residuo solido è composto da sostanze base, al contrario l'idronio è un acido, essendo anche legato al calcolo del pH. Quindi all'aumentare della concentrazione di parte base, la concentrazione di acido diminuisce. Il test per verificare se la correlazione è significativa ($S=21968$, $p\text{-value}=0.0001$) porta a rifiutare l'ipotesi che la correlazione non sia significativa al 5%, ma non all'1%.